DOCUMENT RESUME

ED 224 959 CE 034 750

AUTHOR Blakemore, Thomas F.; Coker, Charles C.
TITLE The Use of Learning Curves in the Prediction of Vocational Potential: Prediction Error and Accuracy Enhancement Techniques. Research Report.
INSTITUTION Wisconsin Univ.-Stout, Menomonie. Stout Vocational Rehabilitation Inst.
SPONS AGENCY National Inst. of Handicapped Research (ED), Washington, DC.
PUB DATE Jun 82
NOTE 48p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; Diagnostic Tests; *Disabilities; *Evaluation Methods; *Predictive Measurement; *Predictive Validity; Predictor Variables; Situational Tests; Test Validity; Vocational Aptitude; Vocational Education; *Vocational Evaluation; Vocational Rehabilitation; *Work Sample Tests

ABSTRACT

A study was conducted (1) to obtain an estimate of the amount of error in prediction that occurs when static measures are used to interpret work-sample performance; (2) to identify a number of dynamic performance measures that reflect learning that might occur on the task; and (3) to compare the predictive accuracy of those measures to that of static measures. The participants in the study were 10 male and 10 female handicapped vocational evaluation clients. The clients practiced on a work sample involving a relatively simple psychomotor task for 5 consecutive work days (50 trials/day). The data collected showed that the participants improved an average of 31 percent in performance speed over the 5 days, with 11 exceeding the industrial standard by day 5. The accuracy of eight methods of predicting day 5 performance was investigated; all of the dynamic prediction methods proved superior to the traditional static work-sample measure. It was concluded that the use of day 1 total response-time scores seriously underestimates the level of performance that a handicapped individual can potentially achieve on a task following practice, while analyzing the performance data with dynamic performance measures can result in significantly more accurate estimates of the level of performance that someone can achieve. It was suggested that microcomputers be used in data collection and analysis of dynamic performance measures. (KC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED224959

THE USE OF LEARNING CURVES IN THE PREDICTION OF VOCATIONAL POTENTIAL: PREDICTION ERROR AND ACCURACY ENHANCEMENT TECHNIQUES

Research Report

Thomas F. Blakemore

and

Charles C. Coker

June, 1982

Research and Training Center

Stout Vocational Rehabilitation Institute

University of Wisconsin - Stout

Menomonie, WI 54751

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R Fry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

Vocational evaluators frequently use work samples to assess the vocational potential of handicapped individuals. One of the purposes of such assessment is the prediction of specific jobs or job areas where the client would have the greatest likelihood of vocational success. In making those predictions, evaluators often rely upon what can be characterized as static performance measures, such as the mean or total time to complete a task. Those measures are static in that they do not reflect any performance changes that might be occurring during testing. For instance, they fail to reflect any improvement in performance that is occurring due to learning. The present study was conducted in order to obtain an estimate of the amount of error in prediction that occurs when static measures are used to interpret work-sample performance. A second purpose was to identify a number of dynamic performance measures (those which reflect any learning that might occur on the task) and to compare the predictive accuracy of those measures to that of static measures. The third purpose was to use these data as the basis for development of a software package on learning-curve analysis for use on microcomputers. Together, these purposes are directed toward the goal of making learning-curve analysis an easily adoptable and valued tool within vocational evaluation.

The participants in this study were 20 handicapped vocational evaluation clients, 10 males and 10 females. Those individuals practiced on a work sample involving a relatively simple psychomotor task for five consecutive work days (50 trials/day). The latency of each response was automatically recorded by a microcomputer and used in the data analyses.

The results of data analyses indicated that the participants improved an average of 30.68% in performance speed over the five days of practice. Only 1 individual exceeded the industrial standard on the task by Day 1 but 11 did so on Day 5. These findings suggest that using a static measure of Day-1 performance, such as the mean or total time for the session, seriously underestimates the level of performance the individual could attain if given practice at the task because it does not reflect the learning that would occur. No differences in performance were found between males and females on this task.

The accuracy of eight methods of predicting Day-5 performance was investigated. The data from either Day 1 alone or from Days 1 - 4 were used in these analyses. Three measures of predictive accuracy were employed: the degree of correlation between predicted and obtained Day-5 scores, the number of classification errors (incorrectly predicting someone to be above or below standard on Day 5) obtained with each prediction method, and the percentage of error in pre-

dicting the Day-5 scores. The total time taken to complete the Day-1 trials was used as the standard for comparison in these analyses since this measure appears to be the one which is typically used by vocational evaluators when assessing an individual's performance.

All of the dynamic prediction methods proved superior to the traditional static work-sample measure. It was found with all three accuracy measures that predictions derived by fitting the data from Days 1 - 4 to any of six different learning curves produced estimates that were significantly more accurate than were obtained using the Day-1 total scores. Those six learning-curve equations,however, did not differ in accuracy on any of the measures. Fitting data from the first 50 trials (Day 1) to learning curves also produced significantly smaller percentage-of-error scores in predicting Day-5 performance than did the use of the Day-1 total scores as predictors. Interestingly, a relatively simple method -- using the mean of the fastest 20% of the 50 trials performed on Day 1 (the "best 20% method") -- was as accurate in predicting Day-5 performance as any of the learning curves which used data from Days 1 - 4.

It was concluded that the use of Day-1 total response-time scores seriously underestimates the level of performance that a handicapped individual can potentially achieve on a task following practice. Underestimating the performance

capacity of a person could lead to the erroneous conclusion that the individual is incapable of performing a job well enough to pursue it as an occupation. The results also suggested that analyzing the performance data with dynamic performance measures (i.e., learning curves, best 20%, etc.) can result in significantly more accurate estimates of the level of performance that someone can achieve. Further research is needed, however, to determine which of several dynamic performance measures is most accurate over longer prediction intervals and with different work tasks. It was suggested that the use of microcomputers for the purposes of collecting and analyzing performance data could lead to the widespread adoption of dynamic performance measures by vocational evaluators. This could result in an increase in the accuracy of the predictions that are made about client job potential.

ACKNOWLEDGEMENTS

The authors wish to thank the members of the staff of the Vocational Development Center at the University of Wisconsin-Stout for their cooperation and assistance with this project. Special thanks to Cindy Leicht who assisted with the data collection and to Fred Menz and Tom Czerlinsky for their helpful comments on this manuscript.

This project was funded in part by a Research Grant from the National Institute of Handicapped Research, Department of Education, Washington, D.C. 20202.

TABLE OF CONTENTS

Chapter		Page
	ABSTRACT	ii
	ACKNOWLEDGEMENTS	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	viii
	LIST OF FIGURES	ix
I.	INTRODUCTION	1
II.	METHOD	11
	Subjects	11
	Apparatus	11
	Procedure	13
III.	RESULTS	14
	Analysis of Performance Change	14
	Comparison of Prediction Formulas	20
	Accuracy of predictions based on Day 1 - 4 data	20
	Accuracy of predictions using Day-1-only data	24
	Overall Analyses	28
IV.	DISCUSSION	29
	Conclusions	35
	REFERENCES	37

LIST OF TABLES

Table		Page
1	Summary of Analysis of Variance of Total Times and Group Means	17
2	Daily Performance Times (in Minutes) for Each Client on Days 1 and 5	18
3	Summary of Statistics Comparing Accuracy of Different Learning Curves (Using Data from Days 1 - 4) and Day-1 Total-Performance Scores in Predicting Day-5 Obtained Scores	22
4	Summary of Statistics Comparing Accuracy of Different Prediction Methods (Using Data from Trials 1 - 50) and Day-1 Total Scores in Predicting Day-5 Obtained Scores	27

LIST OF FIGURES

Figure		Page
1	Mean total response time on the work sample for each daily practice session (50 trials/session)	15

I. INTRODUCTION

One of the tools which vocational evaluators have traditionally used to determine the vocational potential of handicapped persons is the simulated job or work sample. Neff (1966) described this approach as an "effort to capitalize on the virtues of both psychometric testing and job analysis, while trying to avoid the limitations of the older approaches." Work samples are structured assessment situations that replicate the actual tasks involved in a particular job. For example, a work sample designed to measure someone's capacity to function as a lathe operator would require the individual to operate a lathe under observation. Proponents of the work-sample approach cite the simplicity in interpreting the results as a strong point. If the individual demonstrates competency on a work sample such as a lathe and the worker wants to be a lathe operator, then, it is logical to recommend placement in that vocation. But the process of vocational evaluation and the role of work samples within that process is not always so simple.

Interpretation of work-sample performance becomes more susceptible to error when performance scores are obtained and are used for prediction purposes. Usually a client's performance is measured during a one-time administration of a work sample. The obtained performance measure, such as the total time to complete the work task, is then compared to norms for competitively employed individuals or other stan-

dards. Typically, the evaluator would then convert the score to a percentile rank or some equivalent measure. The latter would then be used as an aid in making decisions about the likelihood of success if the client were trained and/or placed in that job or occupational area.

As McCray (1979) has suggested, the above approach probably leads to reasonably accurate decisions in those cases where the client performs at the 80th percentile or higher when compared to a competitively-employed norm group. Such clients probably will be successful at the job. Similarly, when the client scores at the 20th percentile or lower, it is probable that this approach usually achieves the desired accuracy when classifying the client as unlikely to reach the level of the industrial norm on the job. This approach can become problematic, however, when individuals perform in the intervening percentiles (the 21st to 79th). In those cases, the predictions that are made about the clients in the task area have questionable validity and can lead to possibly damaging service decisions and recommendations. The results of such erroneous decisions would be the exclusion of capable individuals (false-negatives) or the inclusion of those who could not be successful (false-positives) at the task. In either case, the client would lose, having missed an opportunity for success or having experienced failure, perhaps unnecessarily.

Prediction is always associated with some degree of error. The current methodology of work-sample testing in Vocational Evaluation contributes, perhaps needlessly, to that error. The problem is that static measures (i.e., mean time, total time, or the number of pieces produced during a given time on one day of evaluation) are used to predict behavior. Further, even the variability within that fixed period is ignored because an average measure of performance is used (e.g., mean time). The use of such measures is based upon the assumption that performance on tasks is relatively stable or unchanging. Yet there are results from literally hundreds of studies which demonstrate that people improve with practice at virtually any task, particularly those involving the learning of psychomotor skills (e.g., Bilodeau & Bilodeau, 1961; Newell & Rosenbloom, 1981). Those studies show that behavior is not static but dynamic and that change in performance with practice is to be expected with most tasks.

Another source of concern with the traditional work-sample approach is that it often compares the performance of inexperienced individuals to norms for experienced individuals. It seems that increasing numbers of individuals in vocational evaluation have had little or no experience in the areas that work samples are designed to measure. In addition, many individuals, experienced or inexperienced, do not perform well under test conditions. Both of those fac-

tors tend to lead to the underestimation of the actual level of skill the individual could attain following practice on the task. For instance, Dunn (1976), who analyzed data collected by Botterbusch (1974) on college students who performed on the Stout U-bolt Assembly, found that only 15% of the males and 6% of the females met the industrial standard for the task on the first administration of the work sample. Chyatte (1976) found that handicapped individuals were even less likely to reach the level of the industrial standard when first exposed to a number of commercial work sample tasks. Dunn's further analyses revealed that 55% of the males and 42% of the females in his study did meet the industrial standard by the end of their fourth practice session. Thus, these data support the concern about the validity and utility of single, static performance scores for prediction of vocational potential. Also, it seems clear that some individuals cannot perform at the level of industrial standards initially, but could do so with practice.

These concerns have been raised before (Dunn, 1976) and techniques have been offered to increase the accuracy of predictions that are made using work-sample performance scores. A number of people have suggested using some measure of the rate at which the individual is learning the task. Such an approach would take advantage of the well established finding of a curvilinear relationship between performance level and the amount of practice an individual has had. The rate of

improvement on a task is typically very high initially, but becomes smaller as the amount of practice increases. This relationship when graphically depicted is referred to as a learning curve and reflects the dynamic nature of performance changes with practice.

Tillman (1971), who was one of the first to advocate a learning-curve approach, suggested that clients be allowed to practice on work samples until their performance no longer showed improvement. The final level of performance could then be used to determine the suitability of training or placing the client in the job represented by the work sample. Thus, this procedure would allow the people to learn as much as they were capable of prior to making a decision about their capacity at the task. A number of people (e.g., Dunn, 1976) have rejected this approach as impractical, however, since performance continues to improve for many thousands of repetitions with some tasks. For instance, Crossman (1959) found that cigar makers continued to show significant improvements in performance for up to four years (over a million task repetitions).

Dunn (1976) suggested that the use of individualized prediction equations (learning curves) could lead to more accurate estimates of client potential on a task. This approach would involve analyzing data from a relatively small number of performance trials with learning-curve formulas.

The parameter values obtained from those analyses would then be used to make predictions (extrapolations) about the level of performance the client could eventually achieve if given ample practice. Dunn tested his assumption using the data collected by Botterbusch (1974). The total performance scores from the first three days of practice were used to predict the performance level on Day 4, the final session. Dunn found that the final performance level could be predicted with less than 1% error, on the average.

Dunn's (1976) findings clearly suggest that learning curves could potentially be used to obtain highly accurate estimates of the level of performance someone can achieve following practice. There are a number of questions which remain to be answered, however, concerning the practicality and accuracy of using learning curves for predictive purposes. For instance, the learning curve approach is often recommended and used by professionals and practitioners, but not for long. The method soon becomes too time consuming. Accurate recording is essential, the data must be analyzed and interpreted for each individual on each task, and alterations may have to be made to existing work-sample procedures. In addition, evaluators often have difficulty interpreting other than "textbook" learning curves and have concerns about the relationship of such curves to existing norms.

It is also unclear which learning-curve formula would result in the most accurate predictions. A wide number of formulas have been used to describe the performance changes that occur with learning, including the one used by Dunn. Three recent articles have made the argument that hyperbolic ($Y = K(X+C)/(X+C+R)$: Mazur & Hastie, 1978), modified exponential ($Y = A + BC^{X}$: Noble, 1978), and power (geometric) functions ($Y = AX^{B}$: Newell & Rosenbloom, 1981) provide the best description of the learning that occurs on a wide variety of both cognitive and psychomotor tasks. It is clear, on the basis of the findings of those three studies, that the formulas examined there provide reasonably accurate descriptions of learning data. However, it is not clear which formula provides the best description of learning and which could provide the most accurate estimates of practiced performance levels because the appropriate comparisons were not made in those studies.

Present Research.

This study is part of an effort by the Research and Training Center to enhance the utilization of dynamic measures of vocational potential - those which incorporate indices of change with practice - and reduce the amount of error in predictions, recommendations, and decisions about vocational potential. The purposes of the present study were to evaluate the extent of change that occurs in motor behavior in relatively brief periods, estimate the amount of error in

prediction, and compare learning-curve equations. The second phase of the Center's research is the development of computer software (programs), interface equipment, and techniques for utilization of microprocessors to allow learning-curve analyses to become an easily adoptable and valued vocational-evaluation tool. The specific objectives of this study were:

1. To estimate the amount of error associated with a static vocational-evaluation method.

2. To identify prediction methods which generate practiced performance levels based upon initial performance levels.

3. To determine which prediction method(s) is the most accurate and practical.

4. To use these findings to develop a microprocessor based learning/performance analysis system.

The participants in this study were 20 handicapped vocational-evaluation clients. These individuals practiced on a work sample for part of each day for five consecutive work days (50 trials/day). The latency of each trial was automatically recorded by a microcomputer and was used in the data analyses. In those analyses, the data from the early

practice sessions (Day 1 alone or Days 1 - 4) were used to predict the level of performance the clients reached during the final practice session (Day 5).

Eight prediction methods were examined in the study. The first method consisted of the total-time score for the first practice session. This static performance measure was included because it appears to be the traditional method used by evaluators. Six of the prediction methods consisted of learning-curve equations. In addition to the hyperbolic, modified exponential, and geometric formulas described above, three other equations were examined. These included an exponential equation ($Y = AB^X$: Spiegel, 1961), a two parameter hyperbolic equation ($Y = A/X + B$: Lippert, 1976), and a log-linear equation ($Y = A + (B \times \log X)$: Dunn, 1976). All of the equations were chosen because they have previously been used to represent learning data.

The accuracy of what has been named the "best-20% method" was also evaluated. This method was developed by the present authors who sought to find a practical yet accurate prediction technique that was based on the well-demonstrated fact that performance improves with practice. This method consisted of using the mean of the fastest 20% of the trials during the first practice session as the estimate of the individual's final performance level. It was assumed that the client would eventually improve with practice to the

point where his or her average response would equal the mean of the fastest x percentage of trials during the initial session. The value of 20% was chosen because it seemed to be reasonable after examining the data.

II. METHOD

Subjects.

The subjects in this study were 20 handicapped adults (10 male, 10 female) who were undergoing vocational evaluation at the Vocational Development Center (VDC) at the University of Wisconsin-Stout. These individuals represented a variety of handicaps, though none were severly handicapped. They were randomly selected with the restriction that they could not participate if they had a disability which would make it impossible to perform on the work sample that was used during testing. Participation was voluntary and subjects were paid $1.50 for each practice session they completed. The data from four subjects were not used because these subjects did not complete five practice sessions. Four additional subjects were chosen to participate in order to replace the discarded data. Informed consent was obtained from all subjects and they were treated in accordance with the policies and procedures established by the University of Wisconsin-Stout on the treatment of human subjects.

Apparatus.

The subjects in this study performed on the Eye-Hand-Foot (EHF) Coordination work sample (Banks, 1974). This is a relatively simple standardized psychomotor task designed to test the ability of an individual to perform tasks which require coordinated eye, hand, and foot movements. Those abilities are required in a wide variety of

work tasks (e.g., machine operation, piloting, etc,). This work sample requires the individual to attach a bolt to a block of wood in a prescribed manner. Assembly of an EHF unit is accomplished using a power drill that is activated by a foot switch. The results of a motion-time study conducted by Banks indicated that the "industrial standard" (mean time to complete the task by an "average" worker in industry) would be 10.80 minutes per 50 units. A second set of norms, which were developed at the Vocational Development Center using vocational evaluation clients, was also used. The latter norms classified people as "above average" (< 14.66 minutes), "average" (14.66 to 22.66 minutes), or "below average" (> 22.66 minutes) depending upon the amount of time taken to complete 50 items. The location of the items to be assembled (nuts, bolts, & wooden blocks) was reversed for right and left handed people.

A microcomputer manufactured by Ohio Scientific Instruments (Model C4P) was used to collect data in this study. The device was linked, via parallel Input/Output ports, to a remote switch which indicated when a subject completed a trial. The computer was programmed to compute the elapsed time taken to complete each trial (assembly of one EHF unit). This information was stored in the computer memory for later data processing. The remote switch was located in a wooden box which was placed on the floor next to the seated subject during each practice session. The final step in the assembly

process required the subject to drop the completed unit into the box, thus triggering the switch.

Procedure.

Subjects were initially given a standardized explanation and demonstration of the correct procedures to use in assembling the EHF units. They then completed five untimed practice trials under the direction of the experimenter. During those trials, the experimenter pointed out any errors the subjects made and answered any questions about the procedure. The subjects, who were tested individually, began testing immediately after the untimed practice trials. Each subject completed 50 trials per day for 5 consecutive work days during free time while they were clients in evaluation at the VDC. Thus, the study consisted of a repeated-measures design in which all subjects were treated identically. The dependent measure was the response latency for each of the 50 repetitions completed each day.

III. RESULTS

Several data analyses were conducted in addressing the research questions. The raw data consisted of response times for the 250 repetitions (50 trials x 5 days) for each of the 20 clients. In the first analysis, the changes in performance over the five sessions were examined. In the subsequent analysis of prediction accuracy, the amount of change in performance from Day 1 to Day 5 served as the criterion against which the different prediction methods were compared. One set of prediction methods used the data from Days 1 - 4 to predict the Day-5 total score. A smaller set of methods used only Day-1 data to predict Day-5 total scores.

Analysis of Performance Change

The initial set of data analyses was conducted to determine how much performance improved with practice. Response times for the 50 trials on each day were summed to produce daily total response-time scores. Average daily total response-time scores across all subjects are given in Figure 1. The mean performance time for each successive daily practice session became smaller, indicating an obvious improvement in performance. (Note that smaller response times reflect better performance on the task).

A 2 (Sex) by 5 (Practice Sessions) analysis of variance (ANOVA) was computed, using the total daily response-time scores for each subject to determine whether performance sig-

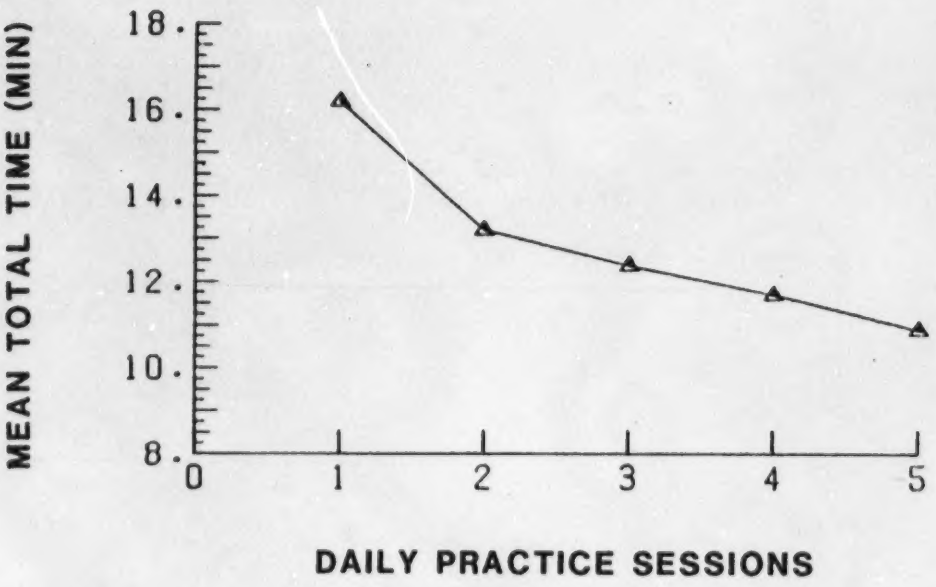

Figure 1. Mean total response time on the work sample for each daily practice session (50 trials/session).

nificantly changed with practice and whether males and females differed at this task. The results of that analysis are presented in Table 1. There was a significant effect for practice only. Post-hoc analyses (Newman-Keuls tests) indicated that performance significantly improved on each succeeding day of practice; i.e., performance on Day 2 was significantly better than on Day 1, performance on Day 3 was better than on Day 2, and so on. The lack of significance for the demographic variable of sex indicates that males and females performed comparably on the task as both groups improved significantly with practice.

The total daily response-time scores from Day 1 and Day 5 were examined to determine the number of subjects who met the industrial standard during each of those days, and to determine the classification of each individual's performance using client norms established at the VDC. Each subject's response times for Days 1 and 5 are listed in Table 2 where it is evident that the performance of every subject improved from Day 1 to Day 5. It can also be seen that only one of the twenty clients (5% of the group) met the industrial standard (a score of 10.80 minutes or less) on Day 1 but that 11 (55%) of the clients did so on Day 5. A chi-square (χ^2) test for related samples (Siegel, 1956) indicated that this change represented a significant increase in the number of individuals who met the industrial standard ($\chi^2(1) = 8.1$, $p < .01$). The clients also showed marked improvement with respect to

TABLE 1

Summary of Analysis of Variance

of Total Times and Group Means

Source	Group Means	df	MS	F	p
Sex (S)		1	8.84	.15	>.10
Male	12.56				
Female	13.16				
Error b		18	58.69		
Practice Sessions (P)		4	79.90	48.33	<.001
Day 1	16.10				
Day 2	13.22				
Day 3	12.42				
Day 4	11.66				
Day 5	10.93				
S x P		4	1.75	1.06	>.10
Error w		72	1.65		

TABLE 2

Daily Performance Times (in Minutes)

for Each Client on Days 1 and 5

Client #	Day 1	Day 5	Client #	Day 1	Day 5
1	12.19[α]	8.30[1α]	11	16.60[β]	12.82[α]
2	18.89[β]	14.65[α]	12	19.92[β]	14.67[β]
3	24.43[C]	12.23[α]	13	15.24[β]	9.62[1α]
4	22.72[C]	14.03[α]	14	13.64[α]	7.14[1α]
5	15.40[C]	9.64[1α]	15	12.37[α]	8.71[1α]
6	13.20[α]	10.19[1α]	16	15.50[β]	11.11[α]
7	16.71[β]	10.16[1α]	17	12.43[α]	8.57[1α]
8	15.03[β]	11.84[α]	18	10.42[1α]	9.53[1α]
9	15.29[β]	11.42[α]	19	26.83[C]	15.12[β]
10	12.28[α]	8.19[1α]	20	12.85[α]	10.67[1α]

[1]Exceeded industrial norm (score < 10.80 minutes) for 50 trials.
[α]Classified as above average by VDC norms.
[β]Classified as average by VDC norms.
[C]Classified as below average by VDC norms.

the VDC client norms. On Day 1, eight clients (40%) performed "above average", nine clients (45%) were "average", and three clients (15%) were "below average". By Day 5, however, 18 clients (90%) performed above average and the remaining 2 (10%) performed at the average level. Thus, the amount of error made when classifying people as likely to be successful or unsuccessful following practice is considerable when the prediction is based on a static-performance measure obtained from a single administration of the work sample.

Further analyses of the data were performed in order to determine the amount of improvement that occurred across practice sessions. For each client, the percentage of improvement between Days 1 and 5 was computed using the following formula: % Improvement = 100 * (1 - (Day-1 score/Day-5 score)). The results of those analyses indicated that the mean rate of improvement for the clients was 30.68%, with a standard deviation of 10.42%. The smallest amount of improvement was 14% and the largest was 50%. Had the performance scores for Day 1 been used as the only estimates of the ability of the clients who participated in this study, their actual capacity would have been underestimated by as much as 50%. Note that even this is a conservative estimate of the error since these individuals probably would have continued to improve with additional practice or training on the task beyond Day 5.

Comparison of Prediction Formulas

The purpose of the following analyses was to determine how accurately the level of performance reached on Day 5 could be estimated using a variety of prediction methods. The accuracy of each method was assessed by examining 1) the degree of correlation between predicted and obtained Day-5 scores; 2) the number of classification errors (incorrectly assigning people as above or below industrial standard on Day 5) made using each prediction method; and 3) the percentage of error in prediction resulting from the use of each method. The data used in these analyses consisted of the total response time scores from Days 1 - 4 or the response times scores for the first 50 trials (Day 1).

Accuracy of predictions based on Day 1 - 4 Data. The first measure of accuracy was a comparison of the degree to which six learning-curve equations and the Day-1 total scores (the measure typically used by evaluators) could predict Day-5 total scores. The data used with each learning-curve equation consisted of the total scores from Days 1 - 4 for each subject. Each of the daily scores represented the total response time for the 50 trials the subject completed that day. For each learning curve, a least-squares fit to the data was calculated and the parameter values that resulted were used to estimate the level of performance the subject obtained on Day 5. Those predicted scores were then correlated with the scores the subjects actually obtained on Day

5. The Day-1 total scores were also correlated with the obtained Day-5 scores.

The results of the correlational analyses are presented in Column A of Table 3, where it can be seen that the correlation between the predicted and obtained scores for the six learning-curve equations were higher (*rs* ranging from .927 to .968) than the correlation between the Day-1 total scores and the Day-5 total scores ($r = .804$). A series of t tests for differences between correlations indicated that the correlations obtained using the learning curves were all significantly higher than the correlations between Day-1 total scores and Day-5 scores (all $ts(17) > 3.17$, $p < .01$). No other significant differences were obtained. Although the correlation between the total response-time scores for Day 1 and Day 5 was reasonably high, it was found that the correlation between the Day-1 total scores and the total scores for each succeeding day became smaller. For instance, the correlation between the Day-1 total scores and the Day-2 total scores was .90, whereas the value dropped to .804 by Day 5.

The second measure of accuracy was an analysis of the number of classification errors that were made using each prediction method. A classification error was made when the use of a particular prediction method indicated that a client would exceed the industrial standard for the work sample on Day 5 but did not (a false positive), and when the prediction

TABLE 3

Summary of Statistics Comparing Accuracy of Different Learning Curves (Using Data from Days 1 - 4) and Day-1 Total-Performance Scores in Predicting Day-5 Obtained Scores

Column:	A	B	C
Prediction Methods	Product-Moment Correlations (Pred. vs. Obtained Day 5 Scores)	Number of Classification Errors**	Mean %(s.d.) Prediction Error
Learning Curves:			
$Y=K(\frac{x+c}{x+c+r})$	.968*	1	6.57 (5.74)
$Y=AB^{x}$	.965*	2	10.13 (6.58)
$Y=A + (B \cdot \log X)$	.954*	1	8.30 (5.89)
$Y= \frac{A}{X} + B$	.947*	0	7.88 (6.02)
$Y=AX^{B}$	.93*	1	7.45 (5.64)
$Y= A + BC^{x}$	.927*	1	8.41 (7.29)
Day-1 Total Scores	.804	10	30.68 (10.42)

*Results of a t test for related correlations indicated that this value significantly differed from the value for the Day-1 Total Scores.

**Consisted of false positive and false negative predictions.

indicated that the individual would be below the industrial standard but the client actually exceeded it (a false negative). For each method, the predicted Day-5 score of each subject was compared to the actual Day-5 status of the client (above or below standard) and the prediction was classified as a correct classification or a classification error. The number of classification errors produced by the six learning curves and the Day-1 total scores are presented in Column B of Table 3. As can be seen there, the learning curves were highly efficient in predicting whether clients would be above or below the industrial standard on Day 5. The use of the Day-1 total scores was not as efficient, however, inasmuch as 10 of the 20 clients were misclassified using this method. Tests for the significance of differences between proportions indicated that the Day-1 total scores produced significantly more classification errors than any of the learning curves (all $\underline{z}$s > 2.76, $p < .01$).

A third measure of accuracy analyzed the percentage of error in prediction that was obtained using each of the learning curves and the Day-1 total scores. This measure was obtained using the following formula: % of Error in Prediction = 100 * (1-(Predicted Value/Obtained Value)). The mean percentages of error in prediction (and the standard deviations) are presented in Column C of Table 3. It can be seen there that the mean percentage of error for the learning curves ranged from 6.57% to 10.13%, whereas, the mean error

rate using the Day-1 total scores as predictors was 30.68%. The percentage-of-error data summarized in Table 3 were analyzed with an ANOVA and Newman-Keuls tests and the results indicated that the use of the Day-1 total scores produced significantly higher percentage-of-error scores, on the average, than did the use of any of the learning curves ($F(6,133) = 30.34$, $p < .001$). Thus, the results of the analysis of this measure, as well as the analyses of the correlational and classification measures discussed above, indicate that the learning curves produce more accurate predictions of Day-5 scores than does the use of the Day-1 total scores. As with the previous measures, however, no significant differences in predictive accuracy were found among any of the learning-curve equations using Day 1 - 4 data.

Accuracy of predictions using Day-1-only data. A second set of analyses was conducted which examined the accuracy of predictions made using simply the data from Trials 1 - 50 ("Day 1 only"). Recall that the above learning-curve analyses used total response-time data from Days 1 - 4. The primary reason for using Day-1-only data in the present analyses was to determine how accurately learning curves would be using the data more likely to be available in a traditional work-sample situation. It is important to have such information because it seems likely that learning curves will not gain widespread usage if the amount of data needed to produce accurate predictions is much more than is currently collected

with most work samples. Thus, these analyses were conducted largely to assess the practicality of using learning curves to evaluate work-sample performance.

A second reason for analyzing the accuracy of learning curves using Day-1-only data was to determine whether the predictions based upon the different learning curves might significantly differ in accuracy if the predictions were made over a longer prediction interval (i.e., predicting from Day 1 to Day 5 rather than from Days 1 - 4 to Day 5). As was the case with the analyses of the data from Days 1 - 4, the predictions based upon the Day-1-only data were examined in terms of the correlation between predicted and obtained Day-5 scores, the number of classification errors, and the percentage-of-error-in-prediction measures. Not all of the learning curves used in the analyses discussed above were used in these analyses[1]. The accuracy of the Day-1 total scores was again used as a standard for comparison when evaluating the accuracy of the predictions made using the Day-1-only data.

In addition to examining the accuracy of learning curves using the Day-1-only data, the "best-20% method was also

---

[1]Information on the accuracy of two formulas ($Y = A + BC^X$ and $Y = AX^B$) was not included because the results of the initial analyses indicated that those formulas were less accurate than the Day-1 total scores. Two other formulas ($Y = AB^X$ and $Y = K(x+c/x+c+r)$) were not included because the computer programs which evaluate those formulas could not efficiently handle such a large amount of data. Those programs use an iterative process to estimate the best-fitting parameter values and this method proved to be very time consuming with 50 data values. Since practicality was one of the criteria used to evaluate the different prediction methods, it was decided not to include those two curves in the analyses of the data from Trials 1 - 50.

studied. This measure was included because of its ease of computation. The measure consisted of identifying the fastest 20% of Trials 1 - 50 and computing the mean of those responses. That score was then used as the estimate of the mean of the 50 trials completed on Day 5.

Table 4 presents a summary of the accuracy of the predictions that were made based upon the Day-1-only data and the Day-1 total scores in predicting Day-5 total scores. As can be seen in the table, the correlation between predicted and obtained Day-5 scores ranged from .57 for the log-linear equation to .83 for the "best-20%" method. Comparisons of those r values, using t tests for related correlations, indicated no significant differences between the correlations obtained using the different methods (all $\underline{t}$s < 1.13, $p > .05$). Significant differences were found, however, for both the number of classification errors and the mean percentage of error in prediction measures for the different prediction methods. Analyses (t tests) for differences between proportions indicated that the number of classification errors made using the Day-1 total scores was significantly higher than either the best-20% method or the use of the log-linear curve. Also, with respect to the percentage of error in prediction measure, an ANOVA and post-hoc Newman-Keuls tests indicated that the best-20% method was more accurate than the two learning curves which were more accurate than using the

TABLE 4

Summary of Statistics Comparing Accuracy of Different Prediction Methods (Using Data From Trials 1 - 50) and Day-1 Total Scores in Predicting Day-5 Obtained Scores

Column:	A	B	C
Prediction Method	Product-Moment Correlations	Number of Classification Errors*	Mean % (S.D.) Prediction Error
Best 20% Method	.835	3**	10.93 (7.36)***
Learning Curves:			
$Y = A/X + B$	.73	7	17.65 (14.51)***
$Y = A + (B \cdot \log X)$	.57	4**	21.75 (17.04)***
Day-1 Total Scores	.804	10	30.68 (10.42)

*Classification errors consisted of false positive and false negative predictions.

**Significantly differed from the Day-1 Total Scores.

***Significantly differed from Day-1 Total Scores

Day-1 total scores in predicting Day-5 scores ($F(3,76) = 8.22$, $p < .001$).

Overall analyses. One final ANOVA was computed, which compared the accuracy of the predictions made with learning curves based upon data from Days 1 - 4 and all of the Day-1-only methods (i.e., Day-1 total scores, best-20% method, and learning curves which used data from Trials 1 - 50). Thus, this analysis compared the accuracy of all prediction methods, regardless of the amount or type of data used to make the predictions. The percentage of error in prediction was the only measure used in this analysis. All learning curves were more accurate than simply using the traditional Day-1 total scores in predicting Day-5 scores ($F(9,190) = 14.01$, $p < .001$). The results of Newman-Keuls tests also indicated that the learning curves which used data from Days 1 - 4 and the best-20% method were significantly more accurate than the two learning curves which used Day-1-only data.

IV. DISCUSSION

The results indicated that the handicapped individuals who participated in this study improved dramatically on the work sample. The clients improved an average of almost one third in just five brief practice sessions. This increase in performance following practice was reflected in the larger number of individuals who met the industrial standard on Day 5 as opposed to Day 1. A similar shift toward higher performance ratings was also found with respect to the VDC client norms. These findings suggest that testing an individual on a work sample only once and using the average or total time for that session as an index of performance capacity for that task can seriously underestimate the level of performance that the individual is capable of achieving on the task.

It was found that males and females did not differ in performance on this task. These results are opposite to those of Dunn (1976), who found that males performed better than females on the Stout U-bolt Assembly, at least initially. However, Noble (1978) concluded, after reviewing the literature on psychomotor performance, that though sex differences are found on many motor tasks, no differences are also found on a large number of other motor tasks. Thus, the fact that Dunn found a sex difference with the task used in his study, but that none was found in this study, is not unu-

sual. It seems important, however, that vocational evaluators know whether a sex difference can be expected on any work sample that they use and the exact nature of that difference. For instance, it would be important to know that females do not perform as well as males early in training (testing) on a given task but that they eventually "catch up" to males with practice. Such a situation could result with tasks on which males would normally have more prior experience than females.

The results of the analyses of the three performance measures were very similar. The findings of the correlational analyses indicated that the predictions based upon the data from Days 1 - 4 were more highly correlated with the obtained Day-5 scores than the predictions based upon any of the methods using data from Day 1 only (i.e., Day-1 total scores, the best-20% method, etc.). In terms of the percentage-of-error measure, the results indicated that the use of learning curves with data from Days 1 - 4 and the best-20% method produced the most accurate estimates of Day-5 performance level. It was also found that the use of learning curves which used data from Days 1 - 4, the best 20% method, and one of the learning curves based upon data from Trials 1 - 50 all produced fewer classification errors than the use of the Day-1 total scores. Thus, it was consistently found over the three measures of accuracy that the worst predictor of Day-5 performance was the measure typically used by

evaluators -- the total time score obtained on a one-time administration of the work sample. The predictions made using the data from Days 1 - 4 were more accurate than those made using data from Day 1 only, with the notable exception of the predictions made with the best-20% method. This best-20% method, which used only the data from Trials 1 - 50 on the first day, was as accurate on two of the performance measures as the methods which used data from four days.

Perhaps the most interesting finding with respect to the percentage of error and the classification error measures was that the best-20% method produced predictions which were as accurate as any of the learning-curve predictions based upon the data from Days 1 - 4. This finding is important because the best-20% method uses data that could be obtained in many vocational-evaluation processes. It would be much easier and less time consuming if vocational evaluators needed to collect data from only a single practice session to obtain an accurate estimate of an individual's future performance capacity.

It is encouraging that the best-20% method produced such accurate predictions but there are still a number of questions remaining about the technique. For instance, it is unclear what percentage of trials would be optimal for use in making predictions. The 20% value used in the analyses of the results of the present study was arbitrarily chosen.

Perhaps taking the mean of the best 10% or 15% would have resulted in more accurate predictions. Also, it is hypothesized that a smaller number or percentage of trials would be needed when making longer range predictions (e.g., predicting performance on Day 25) than when making short range predictions, but that assumption has not been tested.

One issue that was only partially resolved in the present study is the question of which learning-curve formula is most accurate. When the data from Day 1 only were used, two of the curves (the log-linear and the 2 parameter hyperbolic) were found to be more accurate on two of the measures than the other curves. When the data from Days 1 - 4 were used, the six curves that were examined were found to produce equally accurate predictions. These findings do not provide persuasive evidence that any particular learning curve should be used as opposed to the others. In fact, if one were to make a recommendation about which prediction method to use, based upon the present findings, the best-20% method would probably be the most reasonable choice. This method was found to be as accurate as any of the learning curves yet is easier to compute and requires data from only one practice session. This conclusion should be tempered, however, by the possiblity that future research might demonstrate deficiencies in the accuracy of the best-20% method. Research should examine performance over a larger number of trials and with a number of different tasks to further test the accuracy of

this method. Despite this caution, the best-20% method can certainly reduce the amount of prediction error when compared to the use of the traditional static performance measure of work-sample performance.

The decision as to whether learning curves or some static measures are used when evaluating work-sample performance should probably depend upon the reason that the work sample is administered. If the interest of the evaluator and the client is in predicting whether the client is capable of becoming successfully employed at a particular job, then learning curves, or at least some measure reflecting performance change with practice, should be used. If, on the other hand, the purpose of administering the work sample is to document whether the client can or can not perform at a given level at this point in time, then the use of a static measure might be appropriate. In most instances it would be appropriate to use both the static and dynamic measures to analyze the work-sample performance of an individual. This would enable the evaluator to determine how well the client is currently doing relative to other people (the norm group) and/or relative to some performance criterion (e.g., the industrial standard), and to estimate how well the individual could potentially do in the future.

A drawback to the use of either learning curves or some other dynamic performance measure is that such measures

require more work on the part of the evaluator. This is because the use of dynamic performance measures requires the evaluator to either record some performance measure on every response or to collect data from more than one practice session. It is the increased work load involved in their use which has probably prevented dynamic performance measures from becoming a widely adopted practice among vocational evaluators. Apparently, the potential increase in accuracy that the use of such measures could lead to is not offset by the increase in time and effort that their use entails. The use of microcomputers could prove to be invaluable in this context. The advantage of using a microcomputer to monitor the client's performance and then analyze the data is that evaluators are not required to do any more work than they currently do. Thus, a microcomputer makes the use of dynamic performance measures more practical and should lead to an increase in the use of such measures.

As was mentioned previously, the Research and Training Center is currently examining the utility of employing microcomputers to collect and analyze data on client work-sample performance. To date, the Center's efforts have primarily focused on the development of computer programs and interfacing equipment. The present study was an initial effort at evaluating the utility of the learning-curve approach using microcomputers. Future efforts will involve a demonstration/evaluation of the system in a number of

rehabilitation facilities. The interest of those efforts will be in determining both the accuracy and reliability of this approach, as well as its practicality.

Conclusions

The results of this study raise questions about the appropriateness of using static measures of work-sample performance when the purpose of the assessment is to estimate someone's capacity to become successfully employed at the task represented by the work sample. This conclusion seems warranted by the finding that the handicapped subjects in this study increased dramatically in performance on the work sample with only five relatively brief practice sessions. This finding clearly suggests that the use of a static performance measure would seriously underestimate the performance level that an individual could attain on many tasks if given ample practice.

This study also examined the utility of using a number of different prediction techniques for the purpose of estimating someone's performance capacity on a work sample. It was found that the traditional static work-sample measures provided consistently worse estimates of the final performance level than did any of the other techniques used in this study. This finding clearly supports the need to use learning curves or other indices reflecting learning for prediction purposes rather than the traditional static measures

such as the mean or total score. In this study, the best-20% method proved to be as accurate as the six learning curve formulas that were examined and it was suggested that this might be the best method to employ when evaluating work-sample performance. The most obvious advantage of this method is its practicality.

Further research is needed to gain additional information about the utility of using learning curves or the best-20% method to make predictions. For instance, it is still not known which method provides the most accurate estimates of performance over long prediction intervals. Research should also be conducted to develop software programs for use on microcomputers. This rapidly advancing technology could lead to an increase in the use of dynamic performance measures for assessing the work-sample performance of handicapped individuals. Hopefully, the vocational predictions about clients will become increasingly accurate as dynamic measures become a regularly used tool of vocational evaluators.

REFERENCES

Banks, J. J. Revised Tomcheck/Brown eye-hand-foot coordination work sample. Menomonie, WI: Materials Development Center, 1974.

Bilodeau, E. A.(Ed.) Acquisition of skill. New York: Academic Press, 1969.

Bilodeau, E. A., & Bilodeau, I. McD. Motor Skills learning. Annual Review of Psychology, 1961, 12, 243-280.

Botterbusch, K. F. Stout U-bolt assembly work sample. Menomonie, WI: Materials Development Center, 1974.

Chyatte, S. The employment potential of the severely disabled young adult. Emory Research and Training Center 1976 Progress Report, 93-117, 1976.

Crossman, D. A theory of the acquisition of skill-speed. Ergonomics, 1959, 2, 153-166.

Dunn, D. Using competitive norms and industrial standards with work samples. Interface 9. Menomonie, WI: Research and Training Center, 1976.

Lippert, S. Accounting for prior practice in skill acquisition studies. International Journal of Production Research, 1976, 14, 285-293.

McCray, P. An Interpretation of VEWAA/CARF Work Sample Standards. Menomonie, WI: Materials Development Center, 1979.

Mazur, J. E., & Hastie, R. Learning as accumulation: A re-examination of the learning curve. Psychological Bulletin 1978, 85, 1275-1301.

Neff, W. S. Problems of work evaluation. Personnel and Guidance Journal, 1966, 44, 682-688.

Newell, A., & Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive Skills and Their Acquisition. Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc., 1981.

Noble, C. E. Age, race, and sex in the learning and performance of psychomotor skills. In R. T. Osborne, C. E. Noble, and N. Weyl (Eds.), Human Variation. New York: Academic Press, 1978.

Siegel, S. Nonparametric statistics: For the behavioral sciences. New York: McGraw-Hill Book Company, 1956.

Spiegel, M. R. Theory and problems of statistics. New York: McGraw-Hill Book Company, 1961.

Tillman, E. The learning curve in work evaluation. Informational Bulletin, 1971, 3, 1-2.